AD-A113 792 PRINCETON UNIV NJ DEPT OF STATISTICS F/G 12/1
SOME EASY 'T'-STATISTICS.(U)
NOV 81 P S HORN DAAG29-79-C-0205
UNCLASSIFIED TR-214-SER-2 ARO-16669.15-M NL

1 OF 1
AD
A113792

END
DATE
FILMED
5-82
DTIC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A113792

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 16669.15-M	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Some Easy "t"-Statistics		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Paul S. Horn		8. CONTRACT OR GRANT NUMBER(s) DAAG29 79 C 0205
9. PERFORMING ORGANIZATION NAME AND ADDRESS Princeton University Princeton, NJ 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE Nov 81
		13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE APR 26 1982 D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

t-statistics
robustness
order statistics
quartiles

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper explores the use of "t"-statistics based on two or four order statistics. The functions of the order statistics which are used to define the "t"-statistics are the hinges. The hinges are approximately the quartiles and are either exact order statistics or the means of two adjacent order statistics. Two "t"-statistics based on the hinges are examined and compared to other t-statistics, including Student's t, using various criteria.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SOME EASY "t"-STATISTICS

by

Paul S. Horn*

Technical Report No. 214, Series 2
Department of Statistics
Princeton University

November 1981

*Paul S. Horn is Visiting Assistant Professor, Department of Statistics, University of Iowa, Iowa City, Iowa 52242. This research was conducted while the author was a graduate student at Princeton University, Department of Statistics, and was supported in part by the U.S. Army Research Office Contract DAAG29-79-C-0205.

SOME EASY "t"-STATISTICS

Paul S. Horn

A B S T R A C T

This paper explores the use of "t"-statistics based on two or four order statistics. The functions of the order statistics which are used to define the "t"-statistics are the hinges. The hinges are approximately the quartiles and are either exact order statistics or the means of two adjacent order statistics. Two "t"-statistics based on the hinges are examined and compared to other t-statistics, including Student's t, using various criteria.

KEY WORDS: t-statistics, robustness, order statistics, quartiles.

1. INTRODUCTION

The estimation of the location of a sample by a point value, or by a confidence interval, has long been of major concern for statisticians. It is well known that, if the underlying distribution of the sample in question is Gaussian with unknown variance, then the intervals based on Student's t statistics are the best that can be achieved. However, when Gaussianity is not the case, as many who work with real data believe, Student's t intervals can be inappropriate. If the underlying distribution of the sample is heavier-tailed than the Gaussian then the Studnet's t intervals tend to be very long, i.e., conservative (e.g., Benjamini 1980). So, if the sample has more values out in the far ends of the tails than would usually be expected from a Gaussian sample of that size, confidence intervals based on Student's t will not be as precise as they could be if they were otherwise based.

A robust estimate performs well in spite of deviations from ideal behavior, particularly if there are some values in the far ends of the tails. There are many robust estimates of location and spread, as can be seen in <u>The Princeton Robustness Study</u> (Bickel et al. 1972), <u>EDA</u> (Tukey 1977), and <u>Data Analysis and Regression</u> (Mosteller and Tukey 1977). In this study alternative "t"-statistics to Student's t will be formed using a variety of location and spread estimates. Some of these "t"-statistics are not only robust, but are also easy to compute.

Once results are collected concerning the "t"-statistics the next problem is how to interpret them. The notion of criteria for quantifying the performances of the resulting intervals is examined. Traditionally, the Expected Confidence Interval Length, or ECIL, has been emphasized. ECIL was used by both Gross and Kafadar because of its intuitive appeal as a measure of performance of a confidence procedure.

In order to examine other criteria, large classes of criteria are defined along the lines of conservatism of approach. ECIL, then being just a member of a class, is no longer viewed as the principal approach, but one of many.

2. EASILY COMPUTED "t"-STATISTICS

2.1 Order Statistics, Depths, and Hinges

Given n data points we can order them so that $x_1 \le x_2 \le \cdots \le x_n$, where x_i will be referred to as the i^{th} order statistic. We will define the depth of an order statistic as the position of the order statistic with respect to either the minimum, x_1, or the maximum, x_n, whichever is closer. Thus, the order statistics, x_i and x_{n-i+1} both have depth equal to i.

A given depth defines two order statistics: that order statistic whose depth is with respect to the minimum and that whose depth is with respect to the maximum. If the given depth is not an integer but a half-integer, then the corresponding order statistic is not a

single ordered data value, but an average of two ordered data values of adjacent depths. So, for example, if we want the two order statistics whose depth is 1 1/2 then the smaller one is $(x_1+x_2)/2$, and the larger one is $(x_{n-1}+x_n)/2$. The now define the hinge-depth as $([(n+1)/2]+1)/2$ where [ ] is the greatest integer function. Note that this defines two hinges; a lower hinge, x_L, say, and an upper hinge, x_U, where $x_L \le x_U$ by construction. The hinge-depths are approximately $n/4$ for large n, so that hinges are similar to quartiles.

Based on these order statistics, x_L and x_U, the natural estimate of location is the mid-hinge, or $(x_L+x_U)/2$. Similarly, a natural estimate of the spread of the data is the hinge-spread, or x_U-x_L. We can now form a "t"-statistic from just two (or four) such order statistics, namely the hinge-t: $(x_L+x_U)/2(x_U-x_L)$.

We can define other "hinge"'s and thus other "hinge"-spreads and "hinge"-t's. Define the hinge(-)-depth as the hinge-depth minus 1/2. This will yield the hinge(-)-t: $(x_{L-}+x_{U-})/2(x_{U-}-x_{L-})$ with the obvious definitions. Now for a given sample size, either the hinges or the hinge(-)'s will be exact order statistics, while the others will be averages of two adjacent order statistics.

Similarly, let us define hinge(--) as those order statistics whose depth is the hinge-depth -1, and the hinge(+) as those order statistics whose depth is the hinge-depth +1/2. So, there are now four "hinges" to be considered, and thus, four "hinge"-spreads.

2.2 Sampling Situations

To assess how well an estimator performs one must examine it under a variety of situations. In this study we shall examine several estimators and "t"-statistics with various underlying distributions. To keep with tradition, one of these distributions will be the standard Gaussian, but since it is probably rare that real data are distributed in this way, the other distributions used will have heavier tails than the Gaussian.

Since we believe that real data tend to be Gaussian in the middle, then the family defined by Z/U^k, where Z has a Gaussian $(0,1)$ distribution, U is uniformly distributed over the unit interval and independent of Z, and $0 \le k \le 1$, is a useful family for the pruposes just discussed. This family will be referred to as the Slash family of distributions (Rogers and Tukey 1972).

I- this study we will look at four members of this family. The first was already mentioned, that is the case $k = 0$, or the standard Gaussian case. If we now think of $k = 1/v$ where $1 \le v < \infty$ then Z/U^k is Gaussian in the middle, but has tails that behave like the Student's t-distribution with v degrees of freedom (Rogers and Tukey 1972). The other three cases that will be used are $v = 1,2$ and 3, yielding distributions that are Gaussian in the middle but have the tail behavior of the t-distribution on 1, 2 and 3 degrees of freedom respectively. These distributions will be referred to as the Slash, Slasq, and Slacu respectively.

2.3 The "Hinge"-spreads

As described previously we have four "hinges": hinge(--), hinge(-), hinge(+), and the hinge itself. Hence there are four possible "hinge"-spreads, each of which could be used as the denominator of a "t"-statistic: hinge(--)-spread, hinge(-)-spread, hinge(+)-spread, and hinge-spread. At this stage we would like to examine these "hinge"-spreads more closely, and, if possible, narrow our focus to the more sensible of these potential denominators. Hopefully, we can focus on two such "hinge"-spreads: one based on two order statistics and one based on four.

If we let $X = x_1, \ldots, x_n$ be a sample of n data points from a particular distribution then we are interested in comparing the "hinge"-spreads using some appropriate measure of performance. A reasonable masure of performance of $hs(X)$ is its coefficient of variation, or $CV(hs(X)) = \sqrt{Var(hs(X))}/E(hs(X))$. This quantity has the desirable property of being dimensionless.

In order to assess the "hinge"-spreads we have generated, for each sample size, 1000 samples from each of the four distributions mentioned earlier: Gaussian, Slacu, Slasq, and Slash. For each situation (distribution, sample size) we compute the coefficient of variation for each "hinge"-spread. Since we are comparing these spreads, we divide each coefficient of variation by the minimum coefficient of variation of the four spreads in each situation. The results are presented in Table 1, where the sample sizes examined were 5 through 10 inclusive and 20.

As was to be expected, hinge(--)-spread is extremely sensitive to the heavy tails of the Slasq and Slash distributions, especially with small sample sizes. Hinge(+)-spread has the worst performance of the "hinge"-spreads in the Gaussian case while not doing much better than the hinge-spread for the heavy-tailed cases. Some people could argue that Slash is too heavy-tailed a distribution to consider. If we drop Slash as a distribution to be considered, the above remarks concerning hinge(--)-spread and hinge(+)-spread are still valid when Slasq is the most heavy-tailed distribution used.

Hinge-spread and hinge(-)-spread came out the best compromises and so will be the focus of the more detailed study to follow. These two "hinge"-spreads have the advantage of allowing us to choose, for each sample size, a "hinge"-spread based on either two or four order statistics.

To simplify matters further define the pivot as either the hinge or hinge(-), whichever is an exact order statistic, and the bi-pivot as that which is not an exact order statistic. Thus, the depth of the pivot will always be an integer, while that of the bi-pivot will always be a half-integer. We can similarly define the pivot-t as the mid-pivot/pivot-spread, and the bi-pivot-t analogously. So, for example, if we have 5 data points, with the hinge depth = 2, the pivot-t would be the hinge-t and the bi-pivot-t would be the hinge(-)-t. If we have 7 data points, with the hinge depth = 2.5, the reverse is true. So for each sample size then there will be a "t"-statistic based on two order statistics and another based on four.

2.4 Five "t"-statistics Examined

In this study five "t"-statistics, each of the form location estimate/spread estimate, and the confidence intervals for location derived from them will be examined and compared. Two of these are the hinge-t and the hinge(-)-t, equivalently the pivot-t and bi-pivot-t, with the correspondence depending on the sample size. The first "t"-statistic that will be compared with these will be the traditional Student's t, $\sqrt{n}\bar{x}/s$, where $\bar{x}$ is the sample mean, s is the sample standard deviation, and n is the sample size.

The other two "t"-statistics examined are of the more robust type than is Student's t. The first such "t"-statistic is based on the median, which will be referred to as median-t, or med-t. Med-t is defined as $\dot{x}/2\cdot\text{MAD}$, where $\dot{x}$ is the sample median and MAD is the median absolute deviation from the median, or $\text{med}|x_i-\dot{x}|$. The final such "t"-statistic is based on the bisquare weight function, which will be called the biweight-t, or biwt-t. Biwt-t is defined as the c-biweight/s_{bi}, where the biweight $x^* = \Sigma w_i x_i/\Sigma w_i$, where

$$w_i = \begin{cases} (1-u_i^2)^2 & \text{if } u_i^2 < 1 \\ 0 & \text{else} \end{cases}$$

$u_i = (x_i-x^*)/cS$, S is an estimate of spread, and c is a constant. Here we will take S equal to the MAD and c = 9. Since the biweight is an iterative procedure and we wish to keep things simple we will only look at one-step biweights starting at the median, i.e.,

the initial x^* will be taken to be $\dot{x}$. So, to be more specific in our notation, the biweight to be examined is a one step $c = 9$ biweight, or a w9-biweight.

The spread estimate, s_{bi}, is defined by
$s_{bi}^2 = \Sigma' (x-\dot{x})^2 (1-u^2)^4/[\Sigma' (1-u^2)][-1+\Sigma' (1-u^2)(1-5u^2)]$ where Σ' indicates summation over those x data points for which $u^2 \leq 1$ (Mosteller and Tukey 1977). Thus, the biweight "t"-statistic will be referred to as the w9-biweight-t or, the w9-biwt-t.

3. THE CONSERVATIVE CONFIDENCE INTERVALS

3.1 Construction of Confidence Intervals

Since, for a given "t"-statistic, a two-sided $100(1-\alpha)\%$ confidence interval has the form (location estimate) $\pm$ (spread estimate)$\cdot$($100(1-\alpha/2)\%$ point of location estimate/spread estimate), the problem of computing a confidence interval boils down to finding the appropriate $100(1-\alpha/2)\%$ point. Thus, for each sampling situation (fixed sample size and distribution) and "t"-statistic the appropriate 97.5% point will be computed. In this study we will concentrate our efforts on two-sided 95% confidence intervals, though percent points for other levels are included in the final section.

In order to compute these percent points we take advantage of the fact that the distributions in question are from the Slash family, i.e., distributions of random variables of the form Z/Y, where Z is distributed as a unit Gaussian and Y is positive and

independent of Z. Thus the tail probabilities of the "t"-statistics can be computed using a Monte Carlo location and scale swindle (cf. Simon 1975). The percent points are then obtained by doing a bisection on these probabilities.

3.2 Conservative Confidence Intervals

For a given sampling situation and "t"-statistic, it is assumed that the user will know everything except the underlying distribution of the data. While we wish to be robust we should not be foolhardy. Hence, in order to be conservative, we will use the maximum percent point across the four distributions in each case. The use of these conservative percent points is to insure that the corresponding intervals are not less than level $100(1-\alpha)\%$.

So, for each sampling situation and "t"-statistic the conservative two-sided 95% confidence interval is as follows: (location estimate for "t" from given sampling situation) ± (spread estimate of "t" from same sample) · (conservative 97.5% point of the "t"-statistic).

4. ASSESSING PERFORMANCE

4.1 Deciding among the "t"'s

We have five "t"-statistics: Student's-t, median-t, w9-biweight-t, hinge(-)-t, and hinge-t. Now that we have their respective conservative 97.5% points we can construct conservative 95% confidence intervals. The question arises: How do we decide

among these five "t"-statistics? Clearly some criterion, or measurement of performance is needed. Once we have these criteria we can measure the performance of each "t" and eventually be able to compare the different "t"'s by some measure of efficiency based on the criteria. This aspect will be discussed later.

Given that our intervals will cover the true value 95% of the time it would be desirable that their lengths be as shor t as possible. Thus it seems logical that a measure of performance of a confidence procedure based on "t" would be some measure of the lengths of the confidence intervals given by that procedure.

We will be concerned with the performance of a confidence procedure in a specific situation. By situation we mean a fixed underlying distribution and sample size upon which each interval is based. So let $L_1,\ldots,L_n$ be a sample of n confidence interval lengths from a procedure based on "t" under a specific situation. Let F_n be the empirical distribution function of the L_i's which we will now take to be ordered. Since all that we required of our criterion in the previous paragraph was that it be some function of the confidence interval lengths we will call a functional C a criterion if $C(F_n)$ is nonconstant.

The class of such functionals is a very borad one because different investigative aims may require different criteria. A very conservative criteria might weigh heavily the longer interval lengths. In the other direction, a criterion could downweight the unusually long interval lengths. Clearly some criteria are more reasonable

than others, just as some investigators are more reasonable than others.

The class of criteria contains a large subclass from which most investigators would be satisfied to choose. This is the class of functionals that yield polynomials in the L_i's. Specifically, if P is a criterion then $P \in P_c$, the class of polynomial criteria, if $P(F_n) = a_1 L_1^{b_1} + \cdots + a_n L_n^{b_n}$.

The class of polynomial criteria contains a very important class of functionals. That is, P_c contains the class of functionals that is location and scale invariant. Thus all of the usual location estimates (mean, median, mid-hinge) are in the class P_c.

It should be noted that P_c contains a lot of other functionals which may seem like reasonable criteria but are not location and scale invariant. A very conservative, perhaps foolish, investigator may use the Expected (length^2) as a criterion (i.e., $C(F_n) = \frac{1}{n} \sum_{i=1}^{n} L_i^2$). This is very conservative because it will heavily penalize a procedure that yields even a slightly long-tailed distribution of interval lengths. Since we are neither overly conservative nor overly risky (e.g., $C(F_n) = \sqrt{L_1} + \cdots + \sqrt{L_n}$), we choose to concentrate on the class of criteria such that $b_1 = \cdots = b_n = 1$. It so happens that this is the class of criteria that are location and scale invariant.

Now that we have narrowed our class of criteria to the subclass containing C, such that $C(F_n) = a_1 L_1 + \cdots + a_n L_n$, we should further ask what choices of the a_i's would yield reasonable criteria. Since all the L_i's are ≥ 0 a criterion that examines gaps between

the L_i's would have little advantage over one that looked at the L_i's themselves. Thus, we will restrict our attention to criteria such that $a_i \geq 0$ for all i. Further, since we will often compare among procedures we might as well require that $\sum_{i=1}^{n} a_i = 1$.

What had started out as a huge class of criteria, albeit many of which were foolish, has now been reduced to the class of functionals that yield convex combinations of the ordered interval lengths. While it would seem that we are restricting ourselves to the subclass of location and scale invariant functionals it must be emphasized that these invariance properties came as a result of the restrictions placed on the b_i's, and not from any a priori desirability for location and scale invariance.

It must also be emphasized that the main purpose of these criteria is not necessarily to estimate the location of the distribution length_. All of the criteria in P_c, say, however unreasonable, measure some aspect of the interval lengths; the location parameter may be just one aspect that the investigator may be interested in. Like the invariance properties, point estimation of the location of the distribution of the confidence interval lengths may be achieved by a criterion, but this is due to the restrictions on the a_i's and b_i's and not necessarily on an a priori desirability to estimate location.

4.2 ECIL

Gross (1973) and Kafadar (1979) used as their criterion for measuring the performance of "t"-statistics the ECIL, or the Expected

Confidence Interval Length. This is the obvious measure of performance with which to examine the behavior of the intervals. As in the previous section, ECIL should be thought of as an assessment of the performance of the confidence interval procedure based on "t", and not as an estimate of the location of the sample of the confidence interval lengths based on that "t".

ECIL does, on the surface, seem a reasonable measure of performance of a confidence procedure, giving equal weight to each interval length. However, while the main purpose of ECIL is not in trying to estimate the location of the distribution of confidence interval lengths, and thus not explicitly calling for a more robust procedure, the spirit of robustness need not be totally ignored.

If we judge the desirability of a confidence procedure by examining its ECIL, we may make decisions because of a particularly violent behavior in the tail of the distribution of the lengths. Surely we want to be aware of whether a confidence procedure yields interval lengths that are drastically skewed toward high values. Yet, we may be over-penalizing a confidence procedure because a small percentage of its interval lengths are unusually long. It would seem then that a reasonable measure of performance could be based on a truncated ECIL: the expected confidence interval length of the smallest $\beta\%$, or the $\beta\%$-ECIL.

4.3 $\beta\%$-ECIL

What are good choices of β in light of the previous discussion? We are trying to maintain a balancing act. We want highly skewed,

heavy-tailed confidence procedures to be noticed, yet, if this undesirable behavior occurs only quite occasionally, we do not wish to penalize an otherwise sound confidence procedure. In light of the first part of the balance, we could be overly risky if we used a β less than, say, 90. For example, if a confidence procedure yields interval lengths whose top 20% are unusually long then we would probably want our criterion to reflect this fact. Yet, we may be willing to let our criterion ignore the top 10%. If we use 90%-ECIL we still have a reasonable handle on how violently the intervals behave while not allowing the tail of their distribution to exert too much influence.

One could argue that since we have been conservative in protecting ourselves vis-à-vis the intervals themselves, why not continue in this vein with rega-d to their assessment, i.e., use ECIL. If we choose we can be even more extreme than ECIL. We could take a weighted average of the interval lengths giving the longer lengths more weight. The question boils down to how much we wish the tail of the distribution of the interval lengths to influence our judgment of the corresponding confidence procedure. Averaging the first 90% of the interval lengths seems one reasonable choice.

4.4 β%-CIL

Another criterion for performance of a confidence procedure would be some measure of how stretch-tailed is the distribution of interval lengths. Precisely, at what point do we have β% of the

interval lengths lying to the left, i.e., we could use the β-percentile of the confidence interval lengths, or β%-CIL. As with β%-ECIL we must keep in mind that we are trying to measure performance based on the behavior of the interval lengths. Again, if β is too small we would be using a criteria that might ignore too much of the tail of the distribution of the interval lengths, while a β that is very large might keep us from using a procedure which is usually pretty good. For the above reasons, as with β%-ECIL, $\beta = 90$, or 90%-CIL appears to be a good criterion for the performance of a confidence procedure.

4.5 90%-ECIL vs. 90%-CIL

It is clear that the 90%-ECIL and the 90%-CIL measure different aspects of the confidence procedure, though both appear to be valid criteria. A logical question is: When do the two criteria differ drastically, in the sense that they disagree as to which of two confidence procedures is preferable? In Figure A, the graphs indicate the underlying densities of interval lengths for C_1, confidence procedure 1, and similarly for C_2, confidence procedure 2. The density for C_1 has a longer tail than that of C_2 so the 90%-CIL criterion will favor the second confidence procedure. But, the 90%-ECIL for C_1 is less than that of C_2, thus implying that the first confidence procedure is better.

The differences need not be so drastic. C_1 may be 85% efficient (in some sense) with respect to C_2 using 90%-CIL and 95%

efficient using 90%-ECIL. Our action in either case would be to report both results. One might object here and vehemently argue for a single measure of performance so as to be able to make statements like: "Procedure C_1 is 78% as efficient as procedure C_2 using criterion (blank)." If one must choose between the two criteria we recommend using 90%-ECIL over 90%-CIL because of situations that might occur as in Figure A where we would not like C_1 penalized so much only because its tail is a bit stretched.

4.6 Tetra-efficiencies

We can define the p%-ECIL (distribution)-(relative) efficiency as the inverse squared ratio of the p%-ECIL for each "t"-statistic to the best such p%-ECIL among all five "t"-statistics. So, for example, in the Gaussian case we have, say, the 90%-ECIL for each "t". We can then define the 90%-ECIL Gaussian efficiency of "t" as:

minimum{Gaussian 90%-ECIL for all "t"'s}/{Gaussian 90%-ECIL for "t"}.

We can repeat this for Slacu, Slasq, and Slash yielding four efficiencies for each "t".

In the previous discussion it was pointed out that given a criterion for measuring performance we would then wish to make statements as to the efficiencies of the confidence procedures. We have five "t"-statistics with their respective conservative 97.5% points. For each "t" we draw samples from each of the four distributions

mentioned earlier. This yields for each distribution and "t"-statistic a sample of confidence interval lengths. Also, there are four criterion- (distribution)- (relative) efficiencies for each "t"-statistic, one for each of the four distributions. As with Kafadar's tri-efficiency we can define the (relative) tetra-efficiency of "t" as the minimum {(distribution)-efficiency of "t"}.

Each (distribution)-efficiency answers the question: "If the underlying distribution were in fact (distribution) how well would "t" do relative to the best of the other four "t"'s? The tetra-efficiency then is a response to the fact that in practice we never know the underlying distribution. Thus, to be conservative, we take for each "t" its minimum (distribution)-efficiency. Exhibits 2 and 3 give the tetra-efficiencies for the β%-ECIL and the β%-CIL criteria respectively, for β = 50,80,90,95,99, and 100. (Notice that Slasq never provides the extreme, and that Slacu does only once, at $n = 20$, β = 100%.)

4.7 Comments

The results indicate that all of the hinge-"t"'s outperform, in terms of tetra-efficiency, Student's t and Median-t for sample sizes between 4 and 20. Aside from drops of tetra-efficiency for hinge-t at sample size = 5 (pivot-t) and hinge(-)-t at sample size = 6 (bi-pivot-t) the hinge-"t"'s stand up well to the biweight-t for sample sizes ≤ 9.

As might be expected, hinge-t and hinge(-)-t outperform each other, by and large, according to whichever is the bi-pivot-t for that particular sample size. Again, the exception is at sample size

= 6 where pivot-t outperforms bi-pivot-t by almost 17%.

It is clear from the above results that the relative performances of these "t"-statistics are intimately related to the size of the sample in question. As previously noted there seems to be some strange behavior among the hinge-"t"'s for sample sizes between 4 and 6, e.g., the large drop in tetra-efficiency for the bi-pivot-t at sample size = 6. This is probably due to the fact that this is the largest sample size for which any of the hinge-"t"'s uses the extreme values of the sample.

In light of previous arguments, one could argue that in using Slash as a fourth distribution for such sample sizes one may be over-penalizing some of the confidence procedures. It could be argued that it is rare that such contamination is found and we may not want to protect against it especially in small sample sizes. If Slash is omitted as a fourth corner for sample sizes 4 through 7, what were tetra-efficiencies are now tri-efficiencies. Exhibits 4 and 5 give the tri-efficiencies for β%-ECIL and β%-CIL respectively. The major differences in these exhibits are the performances of Student's t and the hinge-"t"'s, especially hinge(-)-t.

If one adheres to 90%-ECIL as the preferred criterion then Student's t is the relative best "t"-statistic for sample sizes 4 through 7. If, on the other hand, one uses ECIL or 90%-CIL then the bi-pivot-t statistic is the best performer for sample sizes 4 through 6 (hinge-t for sample size 4, hinge(-)-t for sample sizes 5 and 6).

While the biweight-t is not that good for sample sizes 4 through 6, it does start improving for sample sizes 7 through 9. In this range it appears to be very close to the bi-pivot-t. Their differences in 90%-ECIL tetra-efficiency are 1.9%, 4.7%, and 2.6%, the first two in favor of the biweight-t. It is up to the user who prefers tri-efficiency whether these small numbers warrant the extra effort required to compute the biweight-t statistic.

In the range of sample sizes 10 through 20 the biweight-t statistic shows itself to be the clear winner over the hinge-"t"'s by about 20%. However, it must be emphasized that, considering how easy it is to use the hinge-"t"'s, it is surprising how well they perform relative to the biweight-t. When confidence intervals are required quickly it would not be unreasonable to use the bi-pivot-t for sample sizes as large as 20.

It should be noted that while the three criteria (90%-CIL, 90%-ECIL, and ECIL) give different quantitative results, i.e., different tetra-efficiencies, they give similar qualitative results: the hinge-"t"'s do well for sample sizes <10 and the biweight-t does the best for sample sizes ≥ 10.

4.8 Summarizing the Results

In this section we will often summarize results which omit the Slash as a distribution to be considered. The tri-efficiencies of the previous section excluded the Slash in two ways: from determining the conservative percent point and from determining the efficiency.

In this section we will distinguish between these two ways. When Slash is excluded from determining the conservative percent point we will refer to this situation as one of tri-confidence. When Slash is excluded from determining efficiencies we will refer to the resulting numbers as tri-efficiencies. Similarly, when Slash is included we will refer to tetra-confidence situations and tetra-efficiencies respectively. These distinctions will aid in summarizing the results of the previous sections.

In Table 6 we present summary statistics for the 90%-ECIL criterion. These statistics are the means and medians of the tetra-efficiencies for the "t"-statistics across a variety of sample sizes. Note that the pivot-t and bi-pivot-t are used here instead of the hinge(-)-t and hinge-t. The reason for this is that there is a clear difference between the pivot-t and bi-pivot-t as can be seen in Tables 2 through 5. These tables also show the difference between the hinge(-)-t and hinge-t not to be as clear cut.

From Table 6 the clear winner is the biweight-t which maintains a tetra-efficiency of about 85% for most of the ranges of sample sizes. Coming in second and third are the bi-pivot-t and pivot-t respectively. The bi-pivot-t is about 15% less tetra-efficient than the biweight-t for most of the ranges examined. It is up to the user to decide whether the extra effort in computing the biweight-t statistic is worth the 15% gain in efficiency over the simple bi-pivot-t statistic.

Similarly, the pivot-t statistic is only about 6% less

tetra-efficient than the bi-pivot-t statistic. Again, it is up to the user to decide if this modest gain in tetra-efficiency gotten by using the simple bi-pivot-t is worth it instead of using the even simpler pivot-t.

The above remarks can be seen more clearly if we examine the stem-and-leaf diagrams presented in Table 7. This shows clearly the order of the "t"-statistics in terms of performance for sample sizes between 4 and 20 according to the 90%-ECIL criterion. Biweight-t is the best, followed by bi-pivot-t, closely followed by pivot-t, with Median-t a fair fourth place, and Student's t the worst.

Table 8 gives results when Slash is dropped from consideration either in determining the conservative percent points (tri-confidence) and/or determining efficiency (tri-efficiency). Comparing these results with those of Table 6 we note the great improvement of Student's t when Slash is not used to help determine efficiency. The pivot-t and bi-pivot-t improve (relatively) slightly in tri-confidence situations. The bi-pivot-t improves (relatively) a bit more, though, when tri-efficiencies are used.

We also note that for the larger sample sizes Student's t has a tri-efficiency of about 72%. This puts it in third place with respect to tri-efficiency/tri-confidence behind the biweight-t and bi-pivot-t. If we consider tri-efficiencies for tetra-confidence situations then Student's t ties for second place with the bi-pivot-t.

Table 9 compares tri-efficiencies and tetra-efficiencies for

sample sizes 4 through 7. (Note that no distinction need be made between tri-confidence and tetra-confidence since Slash does not control for conservatism for any of the "t"-statistics in this range.) Student's t is the clear winner with the bi-pivot-t not too far behind. As with the tetra-efficiencies, pivot-t is in third place behind the bi-pivot-t. Now, though, the differences are a bit bigger between these latter two: about a 30% difference in tri-efficiency. Thus, using teh bi-pivot-t over the pivot-t gives a gain of about 17% in tetra-efficiency, but yields a gain of 30% in tri-efficiency. These large differences should warrant the use of the slightly more complicated bi-pivot-t statistic.

So it seems that exclusion of the Slash distribution only really matters to Student's t and the bi-pivot-t. Bi-pivot-t is still very good even when Slash is included, but Student's t does poorly (relatively) when tetra-efficiencies are examined. Also, as stated before, in the determination of the conservative percent points, it does not matter for small sample sizes whether or not Slash is included.

In conclusion we would recommend using the bi-pivot-t for the small sample sizes since it makes a good showing in terms of both tetra- and tri-efficiency. We recommend using the biweight-t for sample size = 4 to 7, if one does not mind the extra work and believes the data may be highly contaminated. We recommend the use of Student's t for such small sample sizes <u>if</u> one is very confident that the data are not extremely stretch-tailed.

5. PIVOT-t AND BI-PIVOT-t TABLES

Table 10 gives conservative percent points for the pivot-t and bi-pivot-t for various p-values. This table uses results from one of the four distributions to compute the conservative value. If one wishes to omit Slash as a distribution to be considered the corresponding conservative values are given in Table 11.

Appendix

The numbers used to compute the tetra-efficiencies of Section 4.6 were the $\beta\%$-ECIL and $b\%$-CIL. These numbers wer not given because they are too numerous. However, these numbers are given along with standard error estimates in Horn (1981). The standard error estimates were computed using half-sampling techniques (Tukey 1980).

Also, the swindle used in computing the percent points (Simon 1975) given in the last section provides standard error estimates. However, these standard error estimates are for the tail probability, not the critical value (since this was computed using a bisection). Estimation of the standard error of the critical value is discussed in Horn (1981). The standard error estimates of the critical values were practically always less than 3% of the critical value. The exception occured for the pivot-t with sample sizes = 5 and 6 where the standard error estimate of the critical value was equal to 20% of the critical value.

The random numbers used were generated on a PDP-11. A double shuffler was used in the algorithm.

Table 1. Coefficients of Variation of "Hinge"-spreads (Divided by the minimum in each row).

ss	"Hinge"-spread hinge(--)	hinge(-)	hinge	hinge(+)	Distribution
5	1.00000	1.00382	1.51976	1.51976	Gaussian
	1.13669	1.00000	1.19620	1.19620	Slacu
	1.42480	1.18698	1.00000	1.00000	Slasq
	2.20960	1.99014	1.00000	1.00000	Slash
6	1.00961	1.00000	1.39442	1.48581	Gaussian
	1.13119	1.00000	1.15038	1.17238	Slacu
	1.43526	1.18440	1.01455	1.00000	Slasq
	4.69311	4.35990	1.08413	1.00000	Slash
7	1.00000	1.32659	1.38094	2.03156	Gaussian
	1.00000	1.02637	1.02740	1.45203	Slacu
	2.21717	1.06927	1.00000	1.29146	Slasq
	4.22831	1.51415	1.25714	1.00000	Slash
8	1.00000	1.24472	1.25355	1.61735	Gaussian
	1.24378	1.01459	1.00000	1.29048	Slacu
	1.60629	1.05921	1.00000	1.17550	Slasq
	6.33126	1.49976	1.24268	1.00000	Slash
9	1.00000	1.00609	1.29445	1.37583	Gaussian
	1.01088	1.00000	1.23282	1.31778	Slacu
	1.08814	1.00000	1.16169	1.18185	Slasq
	2.60375	1.98406	1.13425	1.00000	Slash
10	1.00000	1.01290	1.24138	1.30474	Gaussian
	1.02971	1.00000	1.19578	1.19353	Slacu
	1.11580	1.00000	1.07064	1.05630	Slasq
	2.21142	1.73319	1.11052	1.00000	Slash
20	1.00000	1.11158	1.11814	1.26797	Gaussian
	1.00000	1.09747	1.07473	1.17880	Slacu
	1.00000	1.05074	1.05245	1.16787	Slasq
	1.15923	1.08967	1.00000	1.00068	Slash

Table 2. 5%-ECIL Relative Tetra-efficiencies Distributions Where These Occur:
g = Gaussian, c = Slacu, q = Slasq, h = Slash (following numbers).

		β =					
ss	t-statistic	50	80	90	95	99	100
	Student's t	97.5 c	92.3 h	78.4 h	54.2 h	31.3 h	13.4 h
	Median-t	54.5 g	49.1 g	47.8 g	46.9 g	45.9 g	45.7 g
4	Biweight-t	47.5 g	42.9 g	41.6 g	40.9 g	40.4 g	40.2 g
	Hinge(-)-t	92.1 c	86.8 h	75.2 h	53.5 h	32.2 h	14.5 h
	Hinge-t	98.3 g	97.2 g	91.5 h	69.4 h	44.5 h	21.8 h
	Student's t	99.3 c	88.1 h	81.6 h	71.2 h	44.2 h	24.6 h
	Median-t	43.2 g	29.0 g	26.2 g	24.9 g	23.7 g	23.3 g
5	Biweight-t	38.5 g	33.3 g	32.3 g	31.9 g	31.4 g	31.3 g
	Hinge(-)-t	98.4 g	96.9 g	96.6 g	93.2 h	66.3 h	42.1 h
	Hinge-t	51.7 g	37.4 g	34.2 g	32.5 g	31.2 g	30.8 g
	Student's t	68.4 h	45.2 h	31.5 h	21.6 h	8.0 h	3.0 h
	Median-t	62.3 g	51.0 g	47.8 g	46.2 g	44.7 g	44.3 g
6	Biweight-t	72.7 g	67.8 g	66.7 g	66.1 g	65.6 g	65.5 g
	Hinge(-)-t	70.4 h	53.0 h	41.4 h	31.4 h	14.2 h	6.2 h
	Hinge-t	74.0 g	61.4 g	58.2 g	56.3 g	54.7 g	54.4 g
	Student's t	56.7 h	38.9 h	29.2 h	22.2 h	14.2 h	7.4 h
	Median-t	58.4 g	44.9 g	41.4 g	39.8 g	38.6 g	38.2 g
7	Biweight-t	76.9 g	71.8 g	70.5 g	70.0 g	69.5 g	69.3 g
	Hinge(-)-t	83.7 g	72.0 g	68.8 g	67.3 g	66.0 g	65.5 g
	Hinge-t	85.9 g	72.4 g	68.6 g	66.8 g	65.4 g	64.8 g

Table 2 (continued)

ss	t-statistic	p = 50	80	90	95	99	100
8	Student's t	52.0 h	35.3 h	25.3 h	17.3 h	6.9 h	2.4 h
	Median-t	67.5 g	57.1 g	54.1 g	52.5 g	51.1 g	50.6 g
	Biweight-t	85.1 g	81.6 g	81.0 g	80.6 g	80.3 g	80.2 g
	Hinge(-)-t	82.9 h	75.7 g	73.4 g	72.2 g	67.2 h	60.2 h
	Hinge-t	87.0 g	78.7 g	76.3 g	74.8 g	73.9 g	73.6 g
9	Student's t	54.0 h	34.7 h	24.3 h	16.6 h	6.8 h	2.0 h
	Median-t	60.0 g	48.6 g	45.2 g	43.4 g	41.9 g	41.4 g
	Biweight-t	81.5 g	78.3 g	77.5 g	77.0 g	76.6 g	76.5 g
	Hinge(-)-t	89.4 g	82.4 g	80.1 g	78.9 g	77.8 g	67.6 h
	Hinge-t	75.1 g	61.2 g	57.2 g	55.1 g	53.4 g	52.8 g
10	Student's t	41.2 h	25.5 h	17.5 h	12.1 h	4.8 h	1.4 h
	Median-t	68.1 g	58.2 g	55.5 g	53.9 g	52.5 g	52.1 g
	Biweight-t	90.1 g	87.9 g	87.3 g	87.0 g	86.8 g	86.7 g
	Hinge(-)-t	78.2 h	73.9 h	71.1 h	68.9 h	65.0 h	59.4 h
	Hinge-t	77.8 g	67.5 g	64.9 g	63.2 g	61.7 g	61.2 g
11	Student's t	43.6 h	23.6 h	15.3 h	10.1 h	4.4 h	2.4 h
	Median-t	63.3 g	54.0 g	51.4 g	49.9 g	48.4 g	47.8 g
	Biweight-t	89.1 g	86.8 g	86.2 g	85.8 g	85.4 g	85.3 g
	Hinge(-)-t	81.7 g	72.9 g	70.2 g	68.7 g	67.3 g	66.8 g
	Hinge-t	84.5 g	74.8 g	72.0 g	70.2 g	68.6 g	68.0 g
12	Student's t	36.8 h	21.5 h	15.1 h	10.3 h	4.7 h	2.5 h
	Median-t	67.0 g	55.4 g	52.9 g	51.5 g	50.2 g	49.9 g
	Biweight-t	87.8 g	85.7 g	85.2 g	84.8 g	84.4 g	84.3 g
	Hinge(-)-t	76.3 g	69.7 g	67.8 g	66.7 g	65.6 g	65.3 g
	Hinge-t	88.7 g	79.2 g	76.9 g	75.4 g	74.2 g	73.7 g

Table 2 (continued)

ss	t-statistic	β = 50	80	90	95	99	100
	Student's t	36.6 h	20.0 h	13.3 h	9.2 h	5.5 h	2.9 h
	Median-t	68.1 g	58.5 g	56.0 g	54.6 g	53.1 g	52.6 g
13	Biweight-t	93.1 g	91.1 g	90.7 g	90.4 g	90.1 g	90.0 g
	Hinge(-)-t	82.2 g	75.2 g	73.2 g	72.2 g	71.2 g	70.8 g
	Hinge-t	82.5 g	71.3 g	68.1 g	66.5 g	64.8 g	64.2 g
	Student's t	33.6 h	20.5 h	13.9 h	9.5 h	4.4 h	2.2 h
	Median-t	70.7 g	61.9 g	59.3 g	58.0 g	56.7 g	56.1 g
14	Biweight-t	93.1 g	91.0 g	90.7 g	90.5 g	90.2 g	90.2 g
	Hinge(-)-t	74.5 g	69.3 g	67.8 g	67.0 g	66.2 g	65.8 g
	Hinge-t	79.4 g	70.7 g	68.2 g	66.9 g	65.7 g	65.1 g
	Student's t	33.7 h	19.6 h	13.0 h	9.2 h	3.8 h	0.0 h
	Median-t	67.8 g	59.0 g	56.7 g	55.5 g	54.3 g	53.8 g
15	Biweight-t	92.4 g	91.0 g	90.5 g	90.3 g	90.2 g	90.1 g
	Hinge(-)-t	79.5 g	70.9 g	68.7 g	67.6 g	66.5 g	66.1 g
	Hinge-t	83.4 g	74.0 g	71.5 g	70.3 g	68.9 g	68.4 g
	Student's t	30.0 h	16.2 h	10.9 h	7.8 h	4.0 h	0.9 h
	Median-t	70.5 g	62.3 g	59.5 g	58.0 g	56.6 g	56.2 g
16	Biweight-t	92.7 g	91.3 g	91.0 g	90.9 g	90.6 g	90.5 g
	Hinge(-)-t	72.9 g	66.5 g	64.4 g	63.3 g	62.1 g	61.7 g
	Hinge-t	83.9 g	76.4 g	73.9 g	72.6 g	71.2 g	70.8 g

Table 2 (continued)

ss	t-statistic	β = 50	80	90	95	99	100
	Student's t	29.7 h	16.5 h	11.4 h	7.9 h	3.9 h	1.4 h
	Median-t	65.8 g	57.8 g	55.3 g	54.0 g	52.9 g	52.6 g
17	Biweight-t	93.1 g	91.6 g	91.1 g	90.8 g	90.5 g	90.5 g
	Hinge(-)-t	78.6 g	71.7 g	69.5 g	68.4 g	67.4 g	67.0 g
	Hinge-t	83.8 g	73.6 g	70.4 g	68.7 g	67.3 g	66.8 g
	Student's t	28.8 h	14.1 h	8.9 h	6.3 h	2.8 h	1.2 h
	Median-t	69.1 g	61.6 g	59.4 g	58.2 g	57.0 g	56.6 g
18	Biweight-t	93.8 g	92.3 g	92.0 g	91.9 g	91.8 g	91.7 g
	Hinge(-)-t	73.5 g	68.3 g	66.8 g	66.1 g	65.3 g	65.0 g
	Hinge-t	79.4 g	71.6 g	69.2 g	67.9 g	66.6 g	66.2 g
	Student's t	27.4 h	14.3 h	9.1 h	6.2 h	2.9 h	0.1 h
	Median-t	66.2 g	59.9 g	57.7 g	56.3 g	55.0 g	54.7 g
19	Biweight-t	93.4 g	92.3 g	91.9 g	91.7 g	91.6 g	91.5 g
	Hinge(-)-t	73.8 g	68.2 g	66.1 g	64.8 g	63.8 g	63.5 g
	Hinge-t	81.9 g	75.3 g	73.0 g	71.7 g	70.7 g	70.3 g
	Student's t	25.6 h	13.4 h	9.0 h	6.2 h	2.9 h	1.1 h
	Median-t	68.8 g	62.4 g	60.5 g	59.4 g	58.5 g	58.1 g
20	Biweight-t	94.6 g	93.7 g	93.4 g	93.3 g	93.2 g	93.2 g
	Hinge(-)-t	70.7 g	65.2 g	63.6 g	62.8 g	62.0 g	61.7 g
	Hinge-t	79.3 g	73.6 g	71.7 g	70.7 g	69.7 g	69.4 g

Table 3. β%-CIL Relative Tetra-efficiencies Distributions Where These Occur.

g = Gaussian, c = Slacu, q = Slasq, h = Slash (following numbers)

ss	t-statistic	β = 50	80	90	95	99	100
	Student's t	92.8 h	52.7 h	25.8 h	7.3 h	4.9 h	0.7 h
	Median-t	45.3 g	44.3 g	40.5 g	36.3 g	36.3 g	34.4 g
4	Biweight-t	40.3 g	37.7 g	34.5 g	35.2 g	31.5 g	39.4 g
	Hinge(-)-t	86.3 h	50.3 h	28.6 h	8.2 h	5.5 h	0.8 h
	Hinge-t	94.9 g	66.9 h	37.0 h	13.4 h	9.4 h	1.3 h
	Student's t	94.1 h	63.6 h	37.3 h	17.4 h	2.0 h	0.4 h
	Median-t	25.4 g	17.3 g	15.9 g	14.4 g	12.4 g	11.3 g
5	Biweight-t	29.9 g	28.2 g	27.3 g	26.9 g	26.2 g	22.4 g
	Hinge(-)-t	95.2 g	87.6 h	59.8 h	34.5 h	4.8 h	1.0 h
	Hinge-t	33.5 g	24.8 g	21.2 g	19.3 g	18.6 g	4.8 h
	Student's t	61.6 h	20.1 h	7.1 h	2.5 h	0.1 h	0.0 h
	Median-t	47.6 g	37.0 g	33.1 g	32.3 g	27.6 g	27.8 g
6	Biweight-t	65.2 g	62.0 g	60.1 g	61.6 g	59.5 g	31.1 h
	Hinge(-)-t	70.4 h	27.7 h	14.0 h	5.0 h	0.3 h	0.1 h
	Hinge-t	58.1 g	46.4 g	41.8 g	39.5 g	37.4 g	16.7 h
	Student's t	51.1 h	17.3 h	8.4 h	4.6 h	0.7 h	0.0 h
	Median-t	41.6 g	28.8 g	26.4 g	26.7 g	22.2 g	21.1 g
7	Biweight-t	68.2 g	63.5 g	63.6 g	64.6 g	58.7 g	16.0 g
	Hinge(-)-t	65.4 g	55.7 g	52.5 g	51.3 g	45.9 h	3.2 h
	Hinge-t	66.3 g	55.0 g	49.5 g	50.5 g	41.5 g	6.1 h
	Student's t	41.3 h	15.1 h	5.8 h	1.4 h	0.1 h	0.0 h
	Median-t	54.2 g	42.3 g	38.9 g	37.1 g	33.8 g	37.1 g
8	Biweight-t	80.4 g	77.1 g	75.7 g	75.9 g	78.1 g	68.2 q
	Hinge(-)-t	72.5 g	63.7 g	59.4 h	50.7 h	17.0 h	7.8 h
	Hinge-t	77.3 g	65.0 g	60.9 g	60.9 g	29.3 h	16.7 h

Table 3 (continued)

		β =					
ss	t-statistic	50	80	90	95	99	100
	Student's t	40.0 h	15.5 h	4.1 h	1.6 h	0.1 h	0.0 h
	Median-t	45.0 g	33.4 g	28.7 g	26.8 g	25.2 g	23.8 g
9	Biweight-t	74.8 g	73.5 g	72.9 g	70.7 g	72.3 g	56.9 h
	Hinge(-)-t	80.2 g	69.7 g	66.3 h	58.9 h	27.4 h	17.5 h
	Hinge-t	58.8 g	42.8 g	38.1 g	36.6 g	32.0 g	26.5 g
	Student's t	28.8 h	10.2 h	3.1 h	1.3 h	0.0 h	0.0 h
	Median-t	54.5 g	44.3 g	41.0 g	36.2 g	35.3 g	37.0 g
10	Biweight-t	86.1 g	85.2 g	83.6 g	83.8 g	85.3 g	79.2 g
	Hinge(-)-t	75.0 h	63.9 h	55.3 h	49.1 h	22.9 h	14.7 h
	Hinge-t	63.8 g	54.0 g	49.6 g	44.5 g	43.6 g	36.1 g
	Student's t	28.9 h	7.2 h	2.3 h	0.8 h	0.1 h	0.0 h
	Median-t	48.9 g	40.6 g	37.3 g	33.5 g	26.0 g	26.4 g
11	Biweight-t	84.2 g	83.5 g	81.2 g	80.6 g	82.3 g	86.3 h
	Hinge(-)-t	68.1 g	59.9 g	53.1 g	52.9 g	45.6 g	2.3 h
	Hinge-t	70.2 g	60.2 g	53.6 g	51.1 g	48.8 g	5.9 h
	Student's t	26.2 h	8.0 h	3.1 h	1.0 h	0.1 h	0.0 h
	Median-t	50.7 g	41.4 g	39.4 g	35.4 g	34.5 g	30.4 g
12	Biweight-t	84.1 g	82.5 g	80.9 g	79.0 g	78.0 g	73.7 g
	Hinge(-)-t	65.8 g	58.8 g	56.5 g	52.0 g	37.8 g	28.3 h
	Hinge-t	74.4 g	66.0 g	62.8 g	59.4 g	56.8 g	43.9 g

Table 3 (continued)

ss	t-statistic	β = 50	80	90	95	99	100
	Student's t	22.2 h	6.7 h	2.2 h	0.9 h	0.1 h	0.0 h
	Median-t	53.6 g	44.9 g	41.4 g	38.4 g	31.1 g	22.1 g
13	Biweight-t	89.8 g	89.1 g	87.7 g	85.2 g	82.3 g	84.2 g
	Hinge(-)-t	72.3 g	64.7 g	61.2 g	59.5 g	51.0 g	27.2 h
	Hinge-t	66.9 g	55.0 g	52.1 g	45.2 g	39.4 g	35.4 g
	Student's t	23.8 h	6.7 h	2.7 h	0.7 h	0.1 h	0.0 h
	Median-t	57.7 g	47.6 g	45.8 g	43.0 g	35.9 g	23.7 g
14	Biweight-t	88.2 g	86.8 g	89.3 g	87.4 g	85.6 g	85.3 g
	Hinge(-)-t	66.8 g	59.7 g	57.9 g	56.9 g	44.9 h	23.7 h
	Hinge-t	65.8 g	56.2 g	55.1 g	52.7 g	44.6 g	21.7 h
	Student's t	22.4 h	5.8 h	2.0 h	0.8 h	0.0 h	0.0 h
	Median-t	54.1 g	47.0 g	43.5 g	42.4 g	33.6 g	36.3 g
15	Biweight-t	88.1 g	83.3 g	88.2 g	83.1 g	85.0 g	76.6 g
	Hinge(-)-t	66.1 g	58.5 g	55.9 g	53.4 g	47.1 g	45.7 g
	Hinge-t	68.6 g	59.6 g	57.1 g	54.4 g	42.3 g	42.0 g
	Student's t	17.6 h	4.7 h	1.9 h	0.6 h	0.1 h	0.0 h
	Median-t	59.0 g	47.8 g	43.9 g	40.9 g	39.0 g	32.6 g
16	Biweight-t	91.1 g	88.9 g	90.1 g	87.6 g	86.3 g	84.1 g
	Hinge(-)-t	64.3 g	53.6 g	52.4 g	49.7 g	44.7 g	40.5 h
	Hinge-t	74.0 g	61.7 g	59.5 g	56.6 g	51.2 g	45.2 g

Table 3 (continued)

ss	t-statistic	β = 50	80	90	95	99	100
	Student's t	17.0 h	4.9 h	1.9 h	0.5 h	0.1 h	0.0 h
	Median-t	55.2 g	44.1 g	41.3 g	39.6 g	38.9 g	33.3 g
17	Biweight-t	91.2 g	88.4 g	89.3 g	85.9 g	88.2 g	84.9 g
	Hinge(-)-t	69.5 g	59.1 g	57.0 g	53.7 g	51.6 g	41.8 g
	Hinge-t	70.1 g	56.3 g	51.5 g	49.4 g	44.4 g	37.6 g
	Student's t	15.4 h	3.7 h	1.4 h	0.5 h	0.0 h	0.0 h
	Median-t	57.4 g	49.2 g	45.7 g	43.3 g	39.0 g	32.7 g
18	Biweight-t	92.4 g	89.8 g	91.1 g	91.0 g	89.3 g	49.0 h
	Hinge(-)-t	66.4 g	58.7 g	57.7 g	56.7 g	52.8 g	46.7 g
	Hinge-t	68.2 g	58.2 g	54.5 g	51.7 g	47.6 g	49.4 g
	Student's t	16.0 h	3.8 h	1.2 h	0.4 h	0.0 h	0.0 h
	Median-t	56.9 g	48.3 g	42.8 g	41.0 g	36.2 g	36.5 g
19	Biweight-t	91.7 g	89.6 g	89.2 g	90.4 g	89.3 g	89.0 g
	Hinge(-)-t	65.9 g	57.5 g	51.8 g	50.1 g	49.9 g	17.4 h
	Hinge-t	72.4 g	62.7 g	57.5 g	56.7 g	54.2 g	46.1 h
	Student's t	14.4 h	3.7 h	1.4 h	0.4 h	0.0 h	0.0 h
	Median-t	58.4 g	51.1 g	48.3 g	46.9 g	43.5 g	28.1 g
20	Biweight-t	92.7 g	92.2 g	90.9 g	92.3 g	90.1 g	80.5 c
	Hinge(-)-t	62.6 g	55.2 g	53.0 g	52.9 g	45.8 g	44.1 h
	Hinge-t	71.1 g	61.5 g	59.8 g	58.0 g	51.8 g	50.0 g

Table 4. β%-ECIL Relative Tri-efficiencies Distributions Where These Occur:

g = Gaussian, c = Clacu, q = Slasq (following numbers)

		β =					
ss	t-statistic	50	80	90	95	99	100
	Student's t	97.5 c	100.0	99.6 q	97.4 q	94.4 q	91.7 q
	Median-t	54.5 g	49.1 g	47.8 g	46.9 g	45.9 g	45.7 g
4	Biweight-t	47.5 g	42.9 g	41.6 g	40.9 g	40.4 g	40.2 g
	Hinge(-)-t	92.1 c	93.3 q	92.7 q	90.9 q	88.4 q	86.6 q
	Hinge-t	98.3 g	97.2 g	97.0 g	96.6 g	96.1 g	95.8 g
	Student's t	99.3 c	98.4 q	97.1 q	95.2 q	91.2 q	87.9 q
	Median-t	43.2 g	29.0 g	26.2 g	24.9 g	23.7 g	23.3 g
5	Biweight-t	38.5 g	33.3 g	32.3 g	31.9 g	31.4 g	31.3 g
	Hinge(-)-t	98.4 g	96.9 g	96.6 g	96.3 g	96.0 g	96.0 g
	Hinge-t	51.7 g	37.4 g	34.2 g	32.5 g	31.2 g	30.8 g
	Student's t	100.0	100.0	99.7 q	97.4 q	94.0 q	91.0 q
	Median-t	62.3 g	51.0 g	47.8 g	46.2 g	44.7 g	44.3 g
6	Biweight-t	72.7 g	67.8 g	66.7 g	66.1 g	65.6 g	65.5 g
	Hinge(-)-t	96.3 g	95.7 g	95.2 g	94.9 g	94.8 g	94.7 g
	Hinge-t	74.0 g	61.4 g	58.2 g	56.3 g	54.7 g	54.4 g
	Student's t	100.0	100.0	98.7 q	95.1 q	88.1 q	73.8 q
	Median-t	58.4 g	44.9 g	41.4 g	39.8 g	38.6 g	38.2 g
7	Biweight-t	76.9 g	71.8 g	70.5 g	70.0 g	69.5 g	69.3 g
	Hinge(-)-t	83.7 g	72.0 g	68.8 g	67.3 g	66.0 g	65.5 g
	Hinge-t	85.9 g	72.4 g	68.6 g	66.8 g	65.4 g	64.8 g

Table 5. β%-CIL Relative Tri-Efficiencies Distributions Where these Occur:

g = Gaussian, c = Slacu, q = Slasq (following numbers)

ss	t-statistic	β = 50	80	90	95	99	100
4	Student's t	100.0	96.8 c	89.2 q	81.9 q	49.5 q	19.2 q
	Median-t	45.3 g	44.3 g	40.5 g	36.3 g	36.3 g	34.4 g
	Biweight-t	40.3 g	37.7 g	34.5 g	35.2 g	31.5 g	39.4 g
	Hinge(-)-t	91.8 c	89.7 c	82.5 q	75.7 q	52.5 q	20.2 q
	Hinge-t	94.9 g	100.0	93.3 g	90.4 g	83.1 q	29.6 q
5	Student's t	100.0	92.8 q	88.7 q	78.4 q	68.7 q	31.1 q
	Median-t	25.4 g	17.3 g	15.9 g	14.4 g	12.4 g	11.3 g
	Biweight-t	29.9 g	28.2 g	27.3 g	26.9 g	26.2 g	22.4 g
	Hinge(-)-t	95.2 g	95.2 c	93.6 g	92.4 g	93.5 g	53.1 q
	Hinge-t	33.5 g	24.8 g	21.2 g	19.3 g	18.6 g	21.4 g
6	Student's t	100.0	97.6 q	80.8 q	68.6 q	35.3 q	10.6 q
	Median-t	47.6 g	37.0 g	33.1 g	32.3 g	27.6 g	27.8 g
	Biweight-t	65.2 g	62.0 g	60.1 g	61.6 g	59.5 g	58.8 g
	Hinge(-)-t	94.9 g	96.1 g	93.0 g	86.4 q	65.2 q	1910 q
	Hinge-t	58.1 g	46.4 g	41.8 g	39.5 g	37.4 g	22.0 q
7	Student's t	100.0	88.8 q	74.8 q	60.6 q	26.6 q	1.0 q
	Median-t	41.6 g	28.8 g	26.4 g	26.7 g	22.2 g	21.1 q
	Biweight-t	68.2 g	63.5 g	63.6 g	64.6 g	58.7 g	16.0 q
	Hinge(-)-t	65.4 g	55.7 g	52.5 g	51.3 g	47.6 g	43.9 q
	Hinge-t	66.3 g	55.0 g	49.5 g	50.5 g	41.5 g	40.6 g

Table 6. Summary of 90%-ECIL Relative Tetra-efficiencies: (for 95% tetra-confidence)

	Range of Sample Size							
	4-20		6-20		8-20		10-20	
t-statistic	mean	median	mean	median	mean	median	mean	median
Student's t	23	15	16	13	14	13	12	13
Median-t	51	55	53	55	55	56	56	56
Biweight-t	79	87	85	90	88	90	89	90
Pivot-t	64	68	66	68	66	68	67	66
Bi-pivot-t	72	72	69	71	72	72	71	72

Table 7. Stem-and-Leaf Diagrams of 90%-ECIL Relative Tetra-efficiencies for Sample Sizes 7-20:

(95% tetra-confidence)

	"t"-statistic				
	Student's t	Median-t	Pivot-t	Bi-pivot-t	Biweight-t
0	899				
1	01333557				
2	954				
3					
4		15			
5		12455667999	7		
6		0	3446788889	6789	
7			003	111233366	07
8				0	1567
9					00011123

(Values at sample size = 7 are underlined.)

Table 8. Summary of 90%-ECIL Relative Efficiencies.

			Range of Sample Size							
			4-20		6-20		8-20		10-20	
-eff. for	-conf.	t-statistic	mean	median	mean	median	mean	median	mean	median
tetra	either	Student's t	23	15	16	13	14	13	12	13
tri	either	Student's t	80	75	78	75	74	72	72	72
either	either	Median-t	51	55	53	55	55	56	56	56
either	either	Biweight-t	79	87	85	90	88	90	89	90
tetra	tetra	Pivot-t	64	68	66	68	66	68	67	66
tri	tetra	Pivot-t	65	(all other entries same as above)						
tetra	tri	Pivot-t	66	70	68	70	69	70	69	70
tri	tri	Pivot-t	67	(all other entries same as above)						
tetra	tetra	Bi-pivot-t	72	72	69	71	72	72	71	72
tri	tetra	Bi-pivot-t	76	73	73	73	72	73	71	72
tetra	tri	Bi-pivot-t	74	76	72	76	75	76	74	76
tri	tri	Bi-pivot-t	79	76	76	77	75	76	75	77

Table 9. Summary of Relative Efficiencies for Sample Sizes 4 through 7*.

	Tri-efficiency		Tetra-efficiency	
t-statistic	mean	median	mean	median
Student's t	98	98	55	55
Median-t	40	44	40	44
Biweight-t	52	53	52	53
Pivot-t	63	63	58	63
Bi-pivot-t	89	95	75	80

*Distinguishing between tri- and tetra-confidence is unnecessary since Slash does not contorl for the conservative percent point for sample sizes 4 through 7.

Table 10. Conservative One-tailed Percent Points, t_p: $P("t" \le t_p) = p$.
(Distributions used = Gaussian, Slacu, Slasq, Slash)

Pivot-t

n	p = .75	.90	.95	.975	.990	.995	.999
4	.320	.477	.553	.738	1.040	1.331	2.312
5	.387	.869	1.370	2.094	3.715	5.805	16.500
6	.298	.531	.759	1.035	1.505	1.962	3.557
7	.262	.451	.550	.720	.978	1.211	1.985
8	.223	.393	.469	.564	.741	.890	1.293
9	.257	.484	.688	.915	1.265	1.575	2.447
10	.216	.400	.523	.668	.878	1.051	1.584
11	.200	.363	.452	.545	.714	.859	1.281
12	.193	.344	.423	.483	.593	.697	.968
13	.208	.389	.497	.608	.792	.945	1.343
14	.189	.348	.437	.525	.661	.776	1.075
15	.172	.318	.399	.466	.586	.685	.945
16	.164	.299	.374	.435	.507	.591	.822
17	.176	.331	.421	.502	.637	.744	1.009
18	.161	.300	.380	.451	.555	.650	.904
19	.156	.288	.361	.423	.502	.575	.761
20	.143	.266	.447	.397	.464	.519	.678

Bi-pivot-t

n	p = .75	.90	.95	.975	.990	.995	.999
4	.372	.575	.831	1.127	1.614	2.079	3.642
5	.301	.466	.550	.717	.968	1.190	1.856
6	.288	.442	.486	.536	.699	.836	1.219
7	.308	.533	.745	.985	1.355	1.686	2.706
8	.244	.432	.548	.707	.939	1.133	1.650
9	.220	.388	.466	.571	.740	.878	1.279
10	.200	.354	.428	.480	.593	.690	.939
11	.219	.404	.513	.658	.867	1.047	1.571
12	.203	.367	.456	.540	.694	.820	1.152
13	.182	.333	.415	.483	.586	.685	.935
14	.176	.316	.389	.445	.516	.599	.805
15	.184	.343	.434	.531	.677	.797	1.122
16	.171	.315	.397	.467	.575	.672	.942
17	.160	.294	.370	.434	.508	.584	.768
18	.150	.275	.344	.399	.465	.527	.695
19	.163	.303	.383	.453	.558	.641	.862
20	.148	.278	.354	.419	.499	.571	.747

Table 11. Conservative One-tailed Percent Points, t_p: $P(\text{"t"} \le t_p) = p$.
(Distributions used - Gaussian, Slacu, Slasq)

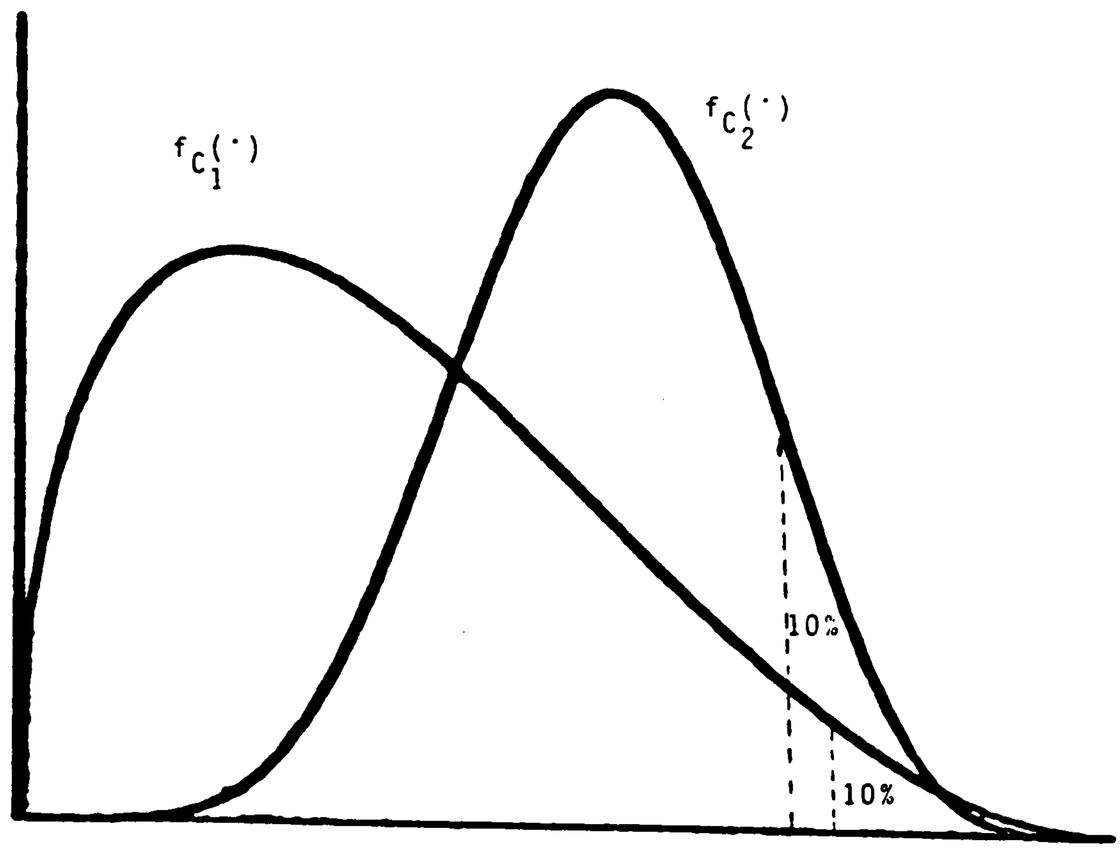
Pivot-t

n	p = .75	.90	.95	.975	.990	.995	.999
4	.238	.428	.553	.738	1.040	1.331	2.312
5	.382	.869	1.370	2.094	3.715	5.805	16.500
6	.263	.531	.759	1.035	1.505	1.962	3.557
7	.211	.407	.550	.720	.978	1.211	1.985
8	.184	.350	.453	.564	.741	.890	1.293
9	.232	.484	.688	.915	1.265	1.575	2.447
10	.195	.384	.523	.668	.878	1.051	1.584
11	.168	.326	.430	.545	.714	.859	1.281
12	.152	.292	.381	.465	.593	.697	.968
13	.184	.363	.486	.608	.792	.945	1.343
14	.165	.321	.425	.525	.661	.776	1.075
15	.151	.291	.379	.465	.586	.685	.945
16	.136	.262	.341	.413	.507	.591	.822
17	.157	.308	.407	.502	.637	.744	1.009
18	.143	.278	.364	.445	.555	.650	.904
19	.133	.256	.333	.405	.502	.575	.761
20	.124	.237	.307	.371	.454	.519	.678

Bi-pivot-t

n	p = .75	.90	.95	.975	.990	.995	.999
4	.303	.575	.831	1.127	1.614	2.079	3.642
5	.229	.425	.550	.717	.968	1.190	1.856
6	.194	.355	.445	.536	.699	.836	1.219
7	.262	.526	.745	.986	1.355	1.686	2.706
8	.212	.413	.548	.707	.939	1.133	1.650
9	.177	.340	.450	.571	.740	.878	1.279
10	.158	.302	.388	.472	.593	.690	.939
11	.194	.381	.513	.658	.867	1.047	1.571
12	.170	.331	.435	.540	.694	.820	1.152
13	.150	.290	.380	.466	.586	.685	.935
14	.138	.264	.343	.415	.516	.599	.805
15	.167	.324	.436	.531	.677	.797	1.122
16	.148	.287	.376	.459	.575	.672	.942
17	.136	.262	.341	.414	.508	.584	.768
18	.126	.242	.315	.381	.465	.527	.695
19	.144	.278	.364	.449	.558	.641	.862
20	.132	.255	.332	.403	.499	.571	.747

Figure A. Underlying Densities of Interval Lengths for Two Confidence Procedures: C_1, C_2



90%-ECIL(C_1) < 90%-ECIL(C_2)
90%-CIL(C_1) > 90%-CIL(C_2)

REFERENCES

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H. and Tukey, J.W. (1972), Robust Estimates of Location: Survey and Advances. Princeton, New Jersey: Princeton University Press.

Benjamini, Y. (1980), The Behavior of the t-Test When the Parent Distribution is Long-Tailed. Ph.D. Dissertation, Department of Statistics, Princeton University, Princeton, New Jersey.

Greenberg, B.G. and Sarhan, A.E. (1962), Contributions to Order Statistics. New York, New York: John Wiley and Sons, Inc.

Gross, A.M. (1973a), A Monte Carlo Swindle for Estimates of Location. Appl. Statist. 22, No. 3, 347-353.

__________ (1973b), Robust Confidence Intervals for the Location of Long-Tailed Symmetric Distributions. Ph.D. Dissertation, Department of Statistics, Princeton University, Princeton, New Jersey.

Horn, P.S. (1981), On Simple Robust Confidence Procedures. Ph.D. Dissertation, Department of Statistics, Princeton University, Princeton, New Jersey.

Huber, P.J. (1964), Robust Estimation of a Location Parameter. Ann. Math. Statist. 35, 73-101.

Kafadar, K. (1979a), A Biweight Approach to the One-Sample Problem. Tech. Report No. 151, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

__________ (1979b), Robust Confidence Intervals for the One- and

Two-Sample Problems. Ph.D. Dissertation, Department of Statistics, Princeton University, Princeton, New Jersey.

Mosteller, F. (1946), On Some Useful 'Inefficient' Statistics. Ann. Math. Statist. 17, 377-407.

__________ and Tukey, J.W. (1977), Data Analysis and Regression: A Second Course in Statistics. Reading, Mass.: Addison-Wesley.

Ralston, A. (1965), A First Course in Numerical Analysis. New York, New York: McGraw-Hill.

Rogers, W.H. and Tukey, J.W. (1972), Understanding Some Long-Tailed Symmetrical Distributions. Statistica Neerlandica, 26, 211-226.

Simon, G. (1975), Swindles to Improve Computer Simulation, With Application of the Problem of Appraising Estimates of Location and Dispersion of Univariate Samples. Tech. Report No. 91, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

Tukey, J.W. (1977), Exploratory Data Analysis. Reading, Mass: Addison-Wesley.

__________ (1980), T^*13: Trendless Jackknifing of Letter Values (and Functions Thereof). Course Notes for Statistics 411, Department of Statistics, Princeton University, Princeton, New Jersey.

FILMED

5-8